

ON MULTIPLE DISCOUNT RATES

CHRISTOPHER P. CHAMBERS

Department of Economics, Georgetown University

FEDERICO ECHENIQUE

Division of the Humanities and Social Sciences, California Institute of Technology

We study the problem of resolving conflicting discount rates via a social choice approach. We introduce several axioms, seeking to capture the tension between allowing for intergenerational comparisons of utility, and imposing intergenerational fairness. Depending on which axioms are judged appropriate, we are led to one of several conclusions: a *utilitarian*, *maxmin*, or a *multi-utilitarian* rule, whereby a utility stream is judged by the worst in a set of utilitarian weighting schemes across discount rates.

KEYWORDS: Global warming, maxmin expected utility, discounting, social discount factor, preference aggregation.

1. INTRODUCTION

A GOVERNMENT ACTOR NEEDS TO CHOOSE a policy with long-lasting, intergenerational consequences, for example, a policy to deal with climate change. She consults with a council of experts, each of whom believes a different discount rate is appropriate for evaluating intertemporal tradeoffs. Since experts' discount rates differ, their policy recommendations differ. Our paper confronts this situation by proposing and evaluating methods of aggregating a decision criterion from a set of experts with different discount rates.

The situation just described is a simplified description of the debate over climate change in economics, at least the debate surrounding the well-known *Stern report*, which has revolved around the problem of choosing a social discount rate.

The climate change problem is a specific instance of a much more general issue highlighted (years before the Stern report) by Weitzman (2001). Weitzman reported the results of a survey of over 2,000 economists. He asked each of them the discount rate that they would use to evaluate long-term projects, such as proposals to abate climate change. The mean of the answers is 3.96% with a standard deviation of 2.94%, reflecting a substantial disagreement over the discount rate. Weitzman considered the possibility that the most prominent economists agree over discounting. So he further ran the survey on a subsample of 50 very distinguished economists (including many who had won, or have since won, a Nobel prize). The results are very similar, with a mean of 4.09% and a standard deviation of 3.07%. It is therefore clear that there is substantial disagreement among economists about the proper discount rate for discounting long-term streams. In fact, Weitzman concluded that disagreements over the discount rate are a critical problem for cost-benefit analysis in general.

Christopher P. Chambers: cc1950@georgetown.edu

Federico Echenique: fede@hss.caltech.edu

We are grateful to Luke Boosey, Simone Cerreia-Vioglio, Vijay Krishna, Michele Le Breton, Efe Ok, Phil Reny, Itai Sher, Tomasz Strzalecki, and participants of numerous seminars and conferences where we have presented the paper, for comments. We wish to especially thank the co-editor and three anonymous referees for suggestions that greatly improved the paper. Finally, Echenique thanks the National Science Foundation for its support through Grants SES 1558757 and CNS 1518941.

The problem of multiple discount rates goes beyond climate change. It shows up in any kind of long-term project evaluation. For example, the U.S. Office of Management and Budget recommends a wide range, between 1% and 7%, for the discount rate when evaluating “intergenerational benefits and costs.” Of course, present-value calculations depend heavily on which number between 1 and 7 is chosen for the discount rate.

What can be done? Imagine a group of experts with different discount rates, and let the set $D \subseteq (0, 1)$ collect their discount rates. How does one decide in the face of wide disagreement over the discount rate? Weitzman proposed the *utilitarian rule*: evaluate a policy according to a weighted sum of the experts’ discount factors in D . Arguably just as reasonable is a *maxmin rule*, where each policy is evaluated by its value according to the expert that values it the least. After all, any one expert could be right, and it makes sense to be conservative in evaluating public policies. As a guide towards choosing a rule, in our paper we take a normative axiomatic approach.

Our approach permits a distillation of the properties that are unique to each rule, and what they have in common. By understanding these properties, we can evaluate utilitarianism and maxmin, and guide the choice of a rule. Moreover, the properties, or axioms, that the utilitarian and maxmin rules have in common give rise to what we call a *multi-utilitarian* criterion. Multi-utilitarianism entails a set of possible utilitarian criteria, each criterion differing in how it weights the discount rates in D . Then, each policy is evaluated according to the utilitarian criterion that values it the least.

The utilitarian and maxmin rules have the property that, whenever all experts rank policy x over y , the rule likewise ranks x over y . This property is a basic unanimity axiom. The government should not contravene a unanimous expert recommendation. We term the axiom D -monotonicity. Note that it is dependent on the exogenous set of discount rates D , which is specified as a primitive of our model.

Aside from D -monotonicity, the utilitarian and maxmin criteria have other properties in common. These properties deal with intergenerational fairness and its prerequisite, intergenerational comparability of utilities. The first property, termed *co-cardinality*, specifies the degree of comparability across different generations’ utilities. It roughly hypothesizes that the utility of different generations is measured in the same units. Co-cardinality is a standard idea in the literature on welfare economics and social choice (it was introduced by d’Aspremont and Gevers (1977) and Sen (1979)). The second property, termed *convexity*, is a notion of intergenerational fairness. Convexity expresses a preference for policies that even out the consequences of policies across the different generations, resulting in “smooth” sequences of generational welfare.

Thus (aside from some technicalities), D -monotonicity, co-cardinality, and convexity are basic normative properties that are shared by the utilitarian and maxmin criteria. We show that they give rise to the multi-utilitarian criterion (Theorem 1). Note that each of these axioms has much going for it. D -monotonicity is the natural requirement that one respects unanimity among experts. Co-cardinality is a standard notion of intergenerational comparability. As with all normative axioms, it may be questionable, but it has a clear meaning and pedigree in welfare economics. Something akin to co-cardinality is simply required in order to meaningfully discuss fairness (Sen (1970)). Finally, convexity is a straightforward fairness requirement. Given the nature of its axioms, multi-utilitarianism is an appealing model.

Now that we understand what utilitarianism and maxmin have in common, and the nature of the model at their intersection, we turn to what is unique about each one of them.

The utilitarian model satisfies a much stronger intergenerational comparability property than co-cardinality. It allows for comparison of changes in intergenerational utility,

but disallows comparisons of absolute levels of utility. A given change in one generation's utility is meaningful regardless of the base utility that generation starts from, but the particular base has no special significance. The axiom is termed *Invariance with respect to individual origins of utilities (IOU)*. IOU is precisely what singles out utilitarianism from multi-utilitarianism (Theorem 2; see the discussion in Section 2.4).

The IOU axiom is more controversial than co-cardinality because it implies that there is no comparable reference point across individual generations' utility. We may want to favor one generation when it is relatively poor, but not when it is well off. This is not allowed under IOU. IOU implies, for example, that if we want to shift utility from generation t to t' in any given situation, then we want to make the shift even after we have added a large amount of utility to generation t' . Co-cardinality would only impose that the utility shift remain desirable after we have added the same utility to all generations. Hence, the IOU criterion is an explicit prohibition of any notion of fairness in absolute levels.

The maxmin model typically violates IOU. Instead, maxmin satisfies a notion of fairness over and above convexity. Fairness commonly dictates that agents' labels, or names, should not matter. In our problem, however, agents are identified by the time-period that corresponds to their generation. We want to allow these to matter. For example, most people want to weight early generations higher than later generations.¹ The property satisfied by the maxmin model is that a particular (not arbitrary) relabeling of agents' names does not matter. The particular relabeling respects the structure of infinite utility streams: it is a "stationary" relabeling; the details are in Sections 2.2 and 2.4. What is important for now is that this new fairness axiom, which we term *invariance to stationary relabeling*, is essentially what distinguishes the maxmin model from multi-utilitarianism (Theorem 3).

Invariance to stationary relabeling is violated by the utilitarian criterion because utilitarianism may value a given intertemporal tradeoff in the near future differently from the same tradeoff in the far future. This issue has been emphasized by Weitzman (2001), who viewed it as a feature, not a bug, and more recently by Zuber (2011) and Jackson and Yariv (2015).

Our results provide a guide to adopting aggregate decisions. Recall our government actor from the beginning of the Introduction. If she follows Weitzman's recommendation to be utilitarian, she implicitly accepts the properties, or axioms, that we show are behind utilitarianism. The same is true of multi-utilitarianism and maxmin. In particular, the choice between utilitarianism and maxmin amounts to choosing between degrees of fairness. Utilitarianism relies on ruling out certain intergenerational comparisons of utility (IOU), comparisons that lie behind fairness considerations. In contrast, maxmin allows for such comparisons, and relies on a stronger notion of intergenerational fairness (IS-TAT). The tradeoff is discussed in detail in Section 2.4. The axioms express basic criteria for how the welfare of different generations are to be compared, and for a degree of intergenerational fairness. Our government actor can decide on which combination of axioms she agrees with, and thus decide on a rule for aggregating experts' recommendations.

¹Certainly most economists seem to prefer to favor early generations over later; of course, disagreement over the extent to which early generations are to be favored is what gives rise to the multiplicity of discount rates. A particularly compelling technical reason for discounting also exists: treating an infinitude of generations identically leads to well-known mathematical impossibilities; see, for example, Svensson (1980), Basu and Mitra (2003), Zame (2007), Lauwers (2010).

2. RESULTS

2.1. *Definitions and Notation*

The objects of choice are sequences, or *streams*, $x = \{x_t\}_{t=0}^{\infty}$ of real numbers, where the natural numbers $\mathbf{N} = \{0, 1, \dots\}$ represent the time line. The indices $t = 0, 1, 2, \dots$ are referred to as *generations*, and the interpretation of x_t is the utility that generation t obtains with stream x . Formulating streams in this way allows us to talk meaningfully about interpersonal comparisons of utility across generations. We assume these interpersonal comparisons are understood and agreed upon by all experts assessing the streams under consideration.

REMARK 1: We could work with a more primitive environment, namely, one in which utils across generations are not primitive, but some more basic object (including lotteries) is the foundation. Such a framework would necessitate a relation \succeq^* operating across pairs (x, t) , where $(x, t) \succeq^* (y, t')$ would be read “generation t is better off with x than t' is with y .” We would then require all experts to agree on \succeq^* .

As a matter of modeling, we need agreement over \succeq^* among the experts so as to focus on the disagreement over the discount rate. In practice, other disagreements will obviously exist, and they may well be important. For the specific case of climate change, the relevant disagreement seems to be about the discount rate, and not about how basic physical outcomes are measured in utils.²

We assume that the space of streams consists of ℓ_{∞} , the set of all bounded sequences, endowed with the norm $\|x\|_{\infty} = \sup\{|x_t| : t \in \mathbf{N}\}$.

When $\theta \in \mathbf{R}$ is a scalar, we often abuse notation and use θ to denote the constant sequence (θ, θ, \dots) . If x is a sequence, we denote by (θ, x) the concatenation of θ and x : the sequence (θ, x) takes the value θ for $t = 0$, and then x_{t-1} for each $t \geq 1$. Similarly, the sequence

$$\underbrace{(\theta, \dots, \theta)}_{T \text{ times}}, x$$

takes the value θ for $t = 0, \dots, T-1$ and x_{t-T} for $t \geq T$.

The notation for inequalities of sequences is: $x \geq y$ if $x_t \geq y_t$ for all $t \in \mathbf{N}$, $x > y$ if $x \geq y$ and $x \neq y$, and $x \gg y$ if $x_t > y_t$ for all $t \in \mathbf{N}$.

2.2. *Axioms*

We suppose a collection of *experts* (distinct from the generations), each of whom is an exponential discounter. The problem is to determine a single decision criterion, a social ranking, modeled as a binary relation \succeq on ℓ_{∞} . All experts agree on how different policies, or plans of action, translate into utility values for each generation. However, they disagree about the discount rate. The set of discount rates possessed by the experts is a closed set $D \subseteq (0, 1)$.

The main substantive axioms we shall consider are as follows (a free variable in a formula should be quantified universally over ℓ_{∞}):

²See, for example, Varian (2006), who summarized the debate in terms of disagreements over discounting, or the survey paper by Nordhaus (2007) for further details.

- *D-monotonicity (D-MON)*: If, for all $\delta \in D$, $(1 - \delta) \sum_t \delta^t x_t \geq (1 - \delta) \sum_t \delta^t y_t$, then $x \succeq y$; and if, for all $\delta \in D$, $(1 - \delta) \sum_t \delta^t x_t > (1 - \delta) \sum_t \delta^t y_t$, then $x \succ y$.
- *Co-cardinality (COC)*: For any scalar $a > 0$ and constant sequence θ , $x \succeq y$ iff $ax + \theta \succeq ay + \theta$.
- *Invariance with respect to individual origins of utilities (IOU)*: For any z , $x \succeq y$ iff $x + z \succeq y + z$.
- *Convexity (CVX)*: For any x, y , constant stream θ , and $\alpha \in [0, 1]$, if $x \succeq \theta$ and $y \succeq \theta$, then $\alpha x + (1 - \alpha)y \succeq \theta$.
- *Invariance to stationary relabeling (ISTAT)*: For all $t \in \mathbb{N}$ and all $\lambda \in [0, 1]$,

$$x \sim \theta \implies \lambda x + (1 - \lambda)(\underbrace{\theta, \dots, \theta}_{t \text{ times}}, x) \sim \theta.$$

In Section 2.4, we interpret and discuss these axioms at length.

We also use two technical axioms:

- *Continuity (CONT)*: $\{y \in \ell_\infty : y \succeq x\}$ and $\{y \in \ell_\infty : x \succeq y\}$ are closed in ℓ_∞ ; and for all $x \in \ell_\infty$ and all θ , if $\theta \succeq (x_0, \dots, x_T, 0, \dots)$ for all T , then $\theta \succeq x$.
- *Compensation (COMP)*: For all t , there are scalars $\bar{\theta}^t$, θ^t , and $\underline{\theta}^t$, with $\bar{\theta}^t > \theta^t > \underline{\theta}^t$, such that

$$(\underbrace{\underline{\theta}^t, \dots, \underline{\theta}^t}_{t \text{ times}}, \bar{\theta}^t, \dots) \succeq \theta^t.$$

There is not a lot to say about the technical axioms CONT and COMP. CONT plays two roles: one is the usual continuity requirement that ensures the existence of a utility representation and the other is as the “monotone continuity” (Villegas (1964), Arrow (1971)) requirement that ensures countably additive priors. Here it can be thought of as continuity at infinity, a basic notion of discounting the future.

COMP ensures that the time horizon truly is infinite. COMP says that there is never a generation beyond which we cease to care about the future.

2.3. Results

Our first result is that a multi-utilitarian criterion emerges essentially from D-MON, COC, and CVX. As we explain in detail in Section 2.4, CVX imposes intergenerational fairness, and COC details how utilities may be compared across generations. Think of multi-utilitarianism as a situation where a decision maker is unsure of the weight to place on each of the different experts. Given a particular weight in the form of $\mu \in \Delta(D)$, her choices are utilitarian, but she evaluates weights in a maxmin fashion.

THEOREM 1: *A weak order \succeq satisfies D-MON, COC, CVX, and CONT iff there is a convex set Σ of Borel probability measures over D such that*

$$U(x) = \min_{\mu \in \Sigma} \sum_{t=0}^{\infty} \left(\int_D (1 - \delta)^t d\mu(\delta) \right) x_t$$

represents \succeq .

REMARK 2: Each expert “agrees” with the axioms in Theorem 1, in the sense that her own behavior satisfies each of these axioms. However, each expert actually agrees with

a stronger set of axioms: namely, each expert agrees that discounting should be exponential. Were we to require the aggregate preference to be exponential, we would run immediately into dictatorship. In light of Arrow's general possibility theorem, this should not be surprising.

REMARK 3: As an anonymous referee has pointed out, the order of integration in Theorem 1 can be reversed here, so that the representation can equivalently be given by

$$U(x) = \min_{\mu \in \Sigma} \int_D \left((1 - \delta) \sum_{t=0}^{\infty} \delta^t x_t \right) \mu(d\delta).$$

This follows from a straightforward application of Tonelli's theorem (see Theorem 11.27 of Aliprantis and Border (1999)). This representation may more readily be interpreted as suggesting a utilitarian criterion.

REMARK 4: Theorem 1 is similar to the main result of Gilboa and Schmeidler (1989), if we interpret the time as states, and consumption streams as uncertain acts. The main difference is the role of D -monotonicity. We leverage classical results relating to the problem of moments; see Shohat and Tamarkin (1943). For our application of the problem of moments, see Lemmas 11 and 12.

The next theorem states that if we add IOU, aggregate choices must be utilitarian.

THEOREM 2: *A weak order \succeq satisfies the axioms in Theorem 1 and IOU iff there is a Borel probability measure μ on D such that*

$$U(x) = \sum_{t=0}^{\infty} \left(\int_D (1 - \delta) \delta^t d\mu(\delta) \right) x_t$$

represents \succeq .

REMARK 5: Actually, Theorem 2 can be proved with a weaker list of properties. CVX is redundant, given the remaining axioms in Theorem 2, as is the full force of COC. We describe this in more detail below, as this is quite relevant for the interpretation of our results. As an anonymous referee has pointed out, the following weakened variant of IOU allows the theorem to be proved yet retains independence of the axioms:

There exists a constant stream θ such that for any x, y satisfying $x \sim y \sim \theta$ and any $\alpha \in [0, 1]$, $\alpha x + (1 - \alpha)y \sim \theta$.

REMARK 6: The measure μ in Theorem 2 is subjective, and dependent on the society under consideration. One can meaningfully ask why there may exist some $\delta \in D$ which is not in the support of μ . This could happen for several reasons. For example, the society may feel that δ is an outlier; or the society could agree that the discount factors in the support of μ already represent a significant "compromise." The explanation for why the support of μ may be smaller than D in our characterization is that D-MON is relatively weak. It claims that if everybody is made weakly better off, then so is society, and if everybody is made strictly better off, then so is society. A stronger concept would require that if everybody is made weakly better off, and at least one individual is made strictly better off, then society should be made strictly better off. Such a notion would force μ to have support equal to D .

Our last result is that if, instead of IOU, we impose ISTAT and the technical COMP requirement, we obtain the maxmin criterion. This is, of course, especially interesting if one finds IOU unpalatable (see the discussion in Section 2.4).

THEOREM 3: \succeq satisfies the axioms in Theorem 1, STAT, and COMP iff there is a nonempty closed set $\hat{D} \subseteq D$ such that

$$U(x) = \min \left\{ (1 - \delta) \sum_{t=0}^{\infty} \delta^t x_t : \delta \in \hat{D} \right\}$$

represents \succeq . Furthermore, \hat{D} is unique.

REMARK 7: Similarly to the preceding, in Theorem 3, \hat{D} need not coincide with D and is itself subjective. The strengthening of D-MON indicated in Theorem 2 would lead to impossibility here, due to CONT. If we were to remove CONT, we would be able to axiomatize some form of lexicographic maxmin solution.

2.4. Discussion of Substantive Axioms

Our substantive axioms stem from two sources. One is the problem of aggregating a single decision criterion from multiple discount rates; the D -monotonicity axiom deals with that. The second source is intergenerational fairness. It should be clear from our discussion of the climate change debate that picking a discount rate amounts to a debate over intergenerational fairness. For example, Varian (2006) expressed that intergenerational fairness is the crucial issue in picking a social discount rate. The COC, IOU, CVX, and ISTAT axioms are about either enabling the discussion of intergenerational fairness, or expressing basic ideas about intergenerational fairness.

2.4.1. D -Monotonicity

D -monotonicity recognizes the existence of a collection of experts who disagree about the discount rate to be used across successive generations. The problem is to come up with a consensus ranking of streams, based on experts' individual judgments. D -monotonicity explicitly models the set of discount rates possessed by these experts. Each $\delta \in D$ is to be interpreted as a discount rate that *some* expert deems possible. D -monotonicity is then a Pareto criterion. If all experts agree that stream x is at least as good as stream y , then the social ranking should also agree on this.

2.4.2. Co-Cardinality and Invariance With Respect to Individual Origins of Utilities

COC and IOU enable a discussion of intergenerational fairness. To talk about intergenerational fairness, one must be able to perform interpersonal comparisons of utility. The reason, as expostulated by Sen (1970), is that, to talk about fairness, one must avoid the conclusion of Arrow's theorem, and Arrow's theorem requires that only individuals' ordinal comparisons matter for aggregate choice. So one has to allow for cardinal comparisons. Sen's proposals were formalized by d'Aspremont and Gevers (1977) into COC and IOU.³ These authors were writing about a general social choice problem; we are interpreting their axioms in an intergenerational setting.

³Specifically Sen's proposals in a paper that was published later as Sen (1979). COC and IOU, and variants thereof, have a long tradition in social choice theory (see the surveys by d'Aspremont and Gevers (2002) or

d'Aspremont and Gevers's COC and IOU axioms can be understood as invariance conditions: if two utility streams are related in some way, then for some class of transformations of streams, the pair under any of the transformations is related in the same way. The larger the class of transformations, the stronger the axiom. Each transformation in the class represents some kind of "information" about the streams which the aggregate ranking cannot detect or utilize. Arrow requires aggregate choices to be blind to everything but individuals' ordinal comparisons, so it cannot allow choices to utilize information about cardinal utility. COC is much more permissive. Yet, it still rules out the use of some types of information. It rules out the use of information about the scale of utils, and it rules out the use of information about common reference points across generations. Roughly, it allows the aggregate to incorporate information about the cardinal structure of utils. COC only asks that the aggregate ranking be blind to the actual common units of measurement of utility utilized, and a common reference point: If, for two streams, all generations' utils are modified by *the same* positive affine transformation, then the aggregate ranking should not change.

IOU, in contrast, disallows information about absolute levels of utility across generations. Thus, aggregate choices must be blind to comparisons of absolute levels of utility. On the other hand, IOU says nothing about differences in utility, which can still be utilized. The informational requirements in IOU are summarized by a different class of transformations. Namely, the addition of a common stream to any given pair of streams should not change the ranking. For example, the decision on the utility gain in going from stream x to y is better for generation t than generation t' is a meaningful statement under IOU, but the claim that generation t is better under x than generation t' is under y is not. We say that comparisons of utility levels are not permitted, but comparisons of utility gains are. (In the words of d'Aspremont and Gevers, "interpersonal comparisons of welfare gains are permitted, while interpersonal comparisons of welfare levels are prohibited.")

It may be instructive to briefly review some examples of when the axioms are violated. A simple example where COC fails is when we believe that there is some "reservation" utility that should be captured by all agents. Say this reservation utility is 0; then a social ranking might recommend $x = (0, 0, 0, \dots) \succ (-1, 10, 10, \dots) = y$ since generation 0 does not reach the reservation under y . However, $y' = (9, 20, 20, \dots) \succ (10, 10, 10, \dots) = x'$ seems a reasonable ranking. So, COC rules out notions of "reservation" utility. The "reversal" comparing x versus y , and then x' versus y' , is based on information about the common level of utility of all generations; observe that x' and y' are obtained from x and y , respectively, by adding 10 utils to all generations. So, COC does not allow us to base aggregate decisions on the common level of utility of all generations and hence rules out models with exogenous "reference points."

Further, COC rules out the use of scale information. For example, suppose that $x = (2, 2, 2, \dots) \succ (3, 1, 3, 1, \dots) = y$ but $y' = (30, 10, 30, 10, \dots) \succ (20, 20, 20, \dots) = x'$. Such a ranking could be made for any number of reasons; but the point is that it is ruled out by COC. Allowing the ranking to depend on scale would be similar to making a decision when utility is measured in dollars, but then changing it when utility is measured in some other units.⁴

Blackorby and Bossert (2008)). The IOU axiom in d'Aspremont and Gevers (1977) also allows for rescaling of utility. Note that IOU bears resemblance to the independence axiom of decision theory.

⁴Our model is silent as to what the unit of measurement of utils is intended to be. It is precisely COC which claims that any positive scale transformation of utils is without loss. Obviously, it has the side effect that, again, reference points or other such objects cannot be given operational meaning.

We demonstrate an example in which IOU is violated. Suppose that we are happy to defer utility from generation 0 to 1 as in $(10, 8, 0, \dots) \succ (14, 4, 0, \dots)$, while $(14, 1004, 0, \dots) \succ (10, 1008, 0, \dots)$. The reason would be that we are happy to defer utility from generation 0 to 1 when 1 is in bad shape, but not when 1 is doing great. This pair of comparisons violates IOU, but not COC. IOU requires an invariance to generation-specific translations of utility, while COC requires invariance to common translations of utility. Put differently, the reversal in the example is based on information about the level of utility of generation 1, and not on the change in utility when we go from one alternative to the other. COC allows decisions to depend on such information, but IOU does not.

IOU also differs from COC in that it introduces an element of separability. For example, $(0, 1, 0, \dots) \succ (0, 0, 2, 0, \dots)$ and $(5, 0, 2, \dots) \succ (5, 1, 0, \dots)$ constitute a violation of IOU because $(5, 0, 2, \dots) - (0, 1, 0, \dots) = (5, 0, 2, \dots) - (0, 0, 2, \dots) = (5, 0, 0, \dots)$. There are circumstances in which such violations of separability are desirable. Imagine that we care about “adjacent” generations because they are filially related. Then we may have $(5, 0, 2, \dots) \succ (5, 1, 0, \dots)$, despite $(0, 1, 0, \dots) \succ (0, 0, 2, 0, \dots)$, because giving 5 to generation 0 and 1 to generation 1 is an unfair advantage to the mother-daughter pair 0 and 1.⁵

2.4.3. Convexity

CVX imposes a basic preference for intergenerational fairness in the form of a preference for “smoothing” utility streams across generations. Such a preference captures fairness considerations rather naturally. For example, the average of the streams $x = (0, 1, 0, 1, \dots)$ and $y = (1, 0, 1, 0, \dots)$ is more fair than either x or y . If we have decided that both x and y are preferable to some constant (and therefore “perfectly smooth”) sequence θ , then the average must be preferable to θ as well. There is not much more to add.⁶

As we mention after Theorem 2, CVX can be dropped from the list of axioms there. Technically, this is because IOU already imposes a relatively strong form of linearity, but this is important for interpretation as well. CVX specifically invokes the idea of “smooth” streams being more fair. At the same time, IOU explicitly eschews reference to comparative absolute levels of utility across generations. It would be strange to impose them jointly; and in fact, Theorem 2 can be read as a statement about what happens when we begin to disregard fairness concerns. The form that arises from the axioms, utilitarianism, is well-understood to lack concern for fairness. We mention here that the full force of COC is also unnecessary for Theorem 2. In fact, COC imposes a type of homogeneity (invariance to scale factor) which can be removed.

2.4.4. Invariance to Stationary Relabeling

ISTAT expresses a different fairness requirement from CVX. ISTAT follows the spirit of *anonymity*, the idea that agents’ labels, or names, should not affect aggregate decisions. This idea is subject to several caveats in this environment.

⁵We should note that this situation is compatible with our maxmin model with $D = \{1/5, 4/5\}$. Given that it violates IOU, it is obviously incompatible with utilitarianism.

⁶CVX is a weakening of the familiar axiom of convexity of preference. Our interpretation of convexity as an intrinsic preference for intertemporal smoothing appears already in Marinacci (1998). Note that, in the standard intertemporal choice model with discounted utility, smoothing is a consequence of the concavity of the utility function. There is no such concavity in our model because the streams under consideration are already measured in “utils” per period of time.

In our problem, agents', or generations', welfare may depend on their names because names are the time period in which they are alive. Because our whole purpose is to study discounting, we want to allow aggregate decisions to depend on the dates of the different generations. Aside from this, as we mentioned in a previous footnote, there are technical reasons to limit this dependence. So, ISTAT limits this dependence. ISTAT says that if a stream x is equivalent to a constant stream θ , and if we treat generation t as we used to treat generation $t - T$, while giving utility θ to generations $0, \dots, t - 1$, then the new stream $(\theta, \dots, \theta, x)$ is also equivalent to θ .

Contrast ISTAT with anonymity. Anonymity says that if a stream x is preferred over y , and x' results from x by relabeling generations, then x' should be preferred over y . Anonymity is not adequate for our problem because many people (and, arguably, most economists) wish to weight earlier generations more heavily than later ones. For example, we may have $(-1, 3, 3, -1, 0, \dots) \sim 0$ and $0 \succ (-1, -1, 3, 3, 0, \dots)$, a violation of anonymity.

In contrast, ISTAT says that $(-1, 3, 3, -1, 0, \dots) \sim 0$ implies $(0, -1, 3, 3, -1, 0, \dots) \sim 0$. Note that $(0, -1, 3, 3, -1, 0, \dots)$ results from $(-1, 3, 3, -1, 0, \dots)$ by treating (or relabeling) generation t as $t - 1$, for $t \geq 1$. Such a relabeling makes sense because time is infinite, *and* because we are comparing with the constant sequence 0. So we can give generation $t = 0$ the utility 0 and generation $t = 0$ is treated the same under the relabeled stream $(0, -1, 3, 3, -1, 0, \dots)$ as under the alternative stream 0. The relabelings we use only make sense when the alternative is a constant stream.

So, ISTAT probably should not be read as a fairness axiom, but rather as an axiom on the consistency of decisions. When a stream is deemed indifferent to a constant, a new stream in which the relative occurrence of the generations' consumption remains the same, but is delayed, should remain indifferent to the constant stream. The notion of delay is that the first generations now consume the value of the constant stream.

More generally, ISTAT says that if $x \sim \theta$, then

$$x' = (\underbrace{\theta, \dots, \theta}_{t \text{ times}}, x) \sim x.$$

Note the resemblance with anonymity: x' results from x by a relabeling of agents' names: $x'_t = x_{t-1}$ ($t \geq T + 1$). This is a relabeling of generations $t = T + 1, T + 2, \dots$. Generations $t = 0, \dots, T$ receive θ , the same as they would receive under the alternative stream θ . In other words: generations $t > T$ are treated in x' the same as generations $t - T$ are treated in x , while generations $t \leq T$ receive the same as under the alternative θ . If $x \sim \theta$, ISTAT requires that $x' \sim \theta$: a particular kind of relabeling of names, one that does not interfere with a preference for earlier generation over later, disciplines how welfare judgments can depend on agents' names.

ISTAT is equivalent to the conjunction of the following two conditions:

$$x \sim \theta \implies (\underbrace{\theta, \dots, \theta}_{t \text{ times}}, x) \sim \theta$$

and

$$[\text{for all } t', \theta \sim x \sim (\underbrace{\theta, \dots, \theta}_{t' \text{ times}}, x)] \implies \lambda x + (1 - \lambda)(\underbrace{\theta, \dots, \theta}_{t \text{ times}}, x) \sim x.$$

We have discussed the first part; we now discuss the second. It requires that whenever there is no benefit or harm from delaying a stream either finitely or indefinitely, then

there can be no hedging or smoothing value to mixing that stream with a delayed version of itself.

Let us compare this with a reinterpretation of the first part of the axiom. The first part can be understood as claiming that, when there is no benefit or loss to delaying a stream *indefinitely*, then there is no benefit or loss to delaying it for any finite amount of time. Roughly, this can be understood as the idea that there are no “smoothing” or hedging opportunities across time between the constant stream θ and x . The second part of the axiom extends this further: the antecedent condition claims that there is no “smoothing” or hedging going on across the relevant streams with respect to time. Since there is no hedging with respect to time, then there should be no benefit to mixing the streams.

Going back to the interpretation of ISTAT as a version of anonymity, suppose that $x \sim \theta$, $x' = (\theta, \dots, \theta, x)$, and $\lambda \in (0, 1)$. If we give generation t a weighted average between what they get in x and in x' , then the stream $\lambda x + (1 - \lambda)x'$ is also indifferent to θ . The meaning is that there is no additional smoothing motive in the comparisons of x with θ , before or after a stationary relabeling. Suppose that \succeq satisfies CVX; then $x \sim \theta$ would imply that $\lambda x + (1 - \lambda)x' \succeq \theta$. ISTAT rules out that $\lambda x + (1 - \lambda)x' > \theta$, which would imply the existence of an additional smoothing motive.

Utilitarianism generally violates ISTAT.⁷ For example, consider $D = \{\frac{1}{100}, \frac{1}{2}\}$ and suppose that the utilitarian criterion places equal weights on the two discount rates in D . Then it turns out that $(-1, 3, 0, \dots) \sim 0.19$ while

$$\underbrace{(0.19, \dots, 0.19)}_{9 \text{ times}}, -1, 3, 0, \dots > 0.19.$$

The reason is that the larger discount rate $(1/2) \in D$ matters more than the smaller rate $(1/100) \in D$ for evaluating tradeoffs far in the future. As a consequence, the utilitarian rule weights the loss of one util by generation 0 the same as a gain of 3 utils by generation 1. For the far-away generations 10 and 11, however, the gain is strictly more important than the loss. This phenomenon was emphasized by [Weitzman \(2001\)](#), [Zuber \(2011\)](#), and [Jackson and Yariv \(2015\)](#).

3. RELATED LITERATURE

Pareto-style properties have a rich history in economics, in particular, Pareto properties for classes of different types of discounters. Classic references include [Böhren and Hansen \(1980\)](#) and [Ekern \(1981\)](#), which characterize Pareto relations for particular classes of discount functions. These papers also take interest in differences of risk attitudes, which we do not discuss. [Trannoy and Karcher \(1999\)](#), [Foster and Mitra \(2003\)](#), and [Bastianello and Chateauneuf \(2016\)](#) provide axiomatizations of (incomplete) Pareto relations in the intertemporal context. A working version of our paper ([Chambers and Echenique \(2016\)](#)) discussed this problem in detail as well.

The concept of multi-utilitarianism mimics that of multiple priors, as in [Gilboa and Schmeidler \(1989\)](#), relying on the natural isomorphism between generations here and states of the world in decision theory. There are works axiomatizing Nash bargaining solutions which can be viewed as maximizing multi-utilitarian welfare functions, for example, [Blackorby, Bossert, and Donaldson \(1994, 1996\)](#), [Ok and Zhou \(2000\)](#), [Hinojosa, Mármol, and Zarzuelo \(2008\)](#). A related concept, also called multi-utilitarianism, appears in [Chambers \(2005\)](#).

⁷It only satisfies it in the case of dictatorship, a point mass on one of the discount rates.

Two recent papers are closely related to our first result (Theorem 1), namely, [Alon and Gayer \(2016\)](#) and [Danan, Gajdos, Hill, and Tallon \(2016\)](#). Interpretation-wise, these papers differ substantially from ours, as they study the aggregation of preferences over state-contingent outcomes. But they explore a facet of the problem of the Bayesian experts (see [Genest and Zidek \(1986\)](#)) which has particularly interested economists. Namely, in an environment in which both “beliefs” and “tastes” can differ, impossibilities arise; classic references are [Hylland and Zeckhauser \(1979\)](#) and [Mongin \(1995\)](#). These impossibilities have been attributed to the notion of spurious unanimity; see [Gilboa, Samet, and Schmeidler \(2004\)](#) and [Mongin \(2016\)](#).

However, viewing time periods as states or vice versa, we see that the mathematical relation to our first result is clear. Each paper considers how to aggregate a finite collection of state-separable preferences.⁸ Mathematically, there is little distinction between “states” and “periods,” so we here stick to the language of time periods, with the understanding that these authors did not restrict to the form of exponential discounting that we do.

One immediate distinction with these prior works is that we assume all experts agree on the interpretation of utils across generations. Both works axiomatize a “set-valued” concept of beliefs closely related to the set of utilitarian weights described in Theorem 1. [Alon and Gayer \(2016\)](#) worked in a framework where social preference satisfies the postulates of [Gilboa and Schmeidler \(1989\)](#), and [Danan et al. \(2016\)](#) assumed social preference conforms to the behavior of [Bewley \(2002\)](#). Motivated by the impossibilities in the problem of the Bayesian experts, and in particular the problem of spurious unanimity, both works weaken the Pareto principle in ways which would not be particularly desirable in our context. Moreover, we do not need to do so, because we assume away disagreement of utils across the experts.

At a technical level, the main distinction is that we allow for an infinite number of possible discount rates in the set D , a possibility afforded by the particular functional form of exponential discounting and well-known polynomial duality techniques, in particular, those of the generalized moment problems surveyed in [Shohat and Tamarkin \(1943\)](#).⁹ Other technical differences are that time periods are infinite (disallowed in [Danan et al. \(2016\)](#)), and exponential discounting is not “non-atomic” (disallowed in [Alon and Gayer \(2016\)](#)).

After Theorem 1, our work diverges significantly. No counterpart of our Theorem 2 appears in these previous works, though it would likely be easy to derive such a result. On the other hand, our Theorem 3 would not even be meaningful in the previous works, and relies on significantly new techniques.

The classic reference uncovering the behavioral implications of additive separable exponential discounting is [Koopmans \(1960\)](#). An implication of Koopmans’s axioms implies what we will call stationarity: roughly, that the passage of time does not affect the ranking of streams, independently of what happens when the time passes. Our weakening of stationarity, ISTAT, is similar but allows for the accommodation of fairness. Roughly, if a stream is indifferent to a constant stream, then delaying the start of the stream should leave society indifferent. In delaying the start of the stream, we imagine that the constant stream is faced until the stream begins. In so doing, we seek to accommodate the intuition of Koopmans’s notion. However, we need to restrict its application to environments in which there is no potential confound with a preference for fairness.

⁸[Danan et al. \(2016\)](#) also studied the more general problem of aggregating multiple priors style preferences.

⁹See also [Zhou \(1997\)](#), who provided an analysis of a representation with an infinite set of agents.

The axiom incorporates some specifics about admissible notions of fairness. First, it presupposes that a constant stream is “maximally fair.” Suppose a stream is indifferent to a status quo. What can happen by delaying the stream? If there are no confounds with fairness, then we would imagine that the delayed stream is also indifferent to the constant stream, since a delay in the status quo simply reproduces the status quo. What could a confound with fairness look like? There are two possibilities. A first possibility is that it would render the delayed stream strictly fairer than the constant stream. Of course, this would not make sense, as the constant stream is maximally fair. Alternatively, the delayed stream might be made strictly less fair than the constant stream. We argue that it is reasonable to suppose that this does not happen. The original stream could not dominate the constant stream across all generations, by monotonicity. So, some generations must do better than the constant stream, and some worse. By replacing the initial individuals’ consumption with the constant stream, we have “smoothed” the stream, giving the initial segment something which is in between the maximal and minimal utility across the following generations. So, we conclude that neither of these two possibilities occur, and confounds with fairness do not arise. The delayed stream is indifferent to the constant stream.

An important motivation for our paper is the literature on multiple discount rates and the evaluation of long-term projects; see [Weitzman \(2001\)](#), [Zuber \(2011\)](#), and [Jackson and Yariv \(2015\)](#). In particular, our focus on the utilitarian rule is motivated by these papers. Jackson and Yariv considered utilitarian aggregation of discounted utilities, and Weitzman argued for the use of an expected discount rate (that he obtained through a survey of economists) like we obtain in Theorem 2. These authors discussed the issue of how the utilitarian rule changes its consideration of intertemporal tradeoffs depending on whether they are in the near or distant future. Weitzman argued that such a change may be desirable. In our context, we capture the issue as a violation of ISTAT (see Section 2.4).

The recent paper by [Feng and Ke \(2017\)](#) deals with the impossibility raised by Jackson and Yariv by weakening their Pareto axiom. Jackson and Yariv required that a social planner respects the rankings of any given generation over future utility streams. Feng and Ke required that the planner respects the ranking of a generation only when all future generations agree on the ranking. By means of this device, they axiomatized the utilitarian rule. Their result can be viewed as an alternative to our Theorem 2, but it relies on significantly different ideas. It assumes that the planner is an exponential discounter, and follows ideas that are close to Harsanyi’s theorem on utilitarianism ([Harsanyi \(1955\)](#)).

Our result on the maxmin model is related to the literature on multiple priors ([Gilboa and Schmeidler \(1989\)](#) and [Huber \(1981\)](#)) by interpreting the set of time periods as a state space. The same approach of identifying time with states was taken by [Marinacci \(1998\)](#) and [Gilboa \(1989\)](#). Marinacci suggested interpreting convexity, or “uncertainty aversion,” as a preference for intertemporal smoothing, as we have done here. [Bastianello and Chateauneuf \(2016\)](#) is a more recent example; they worked out the implications of delay aversion for multiple priors (and other models) representation of intertemporal preferences. The paper by [Wakai \(2008\)](#) also considers a maxmin representation over the discount rate, but in his model the discount rate may be different in each time period. In that sense, his model is closer to Gilboa and Schmeidler’s multiple-prior version of maxmin. Wakai’s focus is on obtaining a dynamically consistent version of the model with multiple and time-varying discount rates.

[Nishimura \(2018\)](#) developed a general theory of intransitive preferences, and introduced the idea of a transitive core. He had an application to time preferences, the primitive being dated consumption, and building on the work on relative discounting of [Ok](#)

and Masatlioglu (2007). Nishimura's Theorem 3 states that the transitive core of a relative discounting preference has a representation with multiple discount functions. His analysis is very different from ours, but has in common the proposal of multiple discount rates for the purpose of making normative welfare comparisons.

4. PROOFS

We shall make use of ℓ_1 , the space of absolutely summable sequences, endowed with the norm $\|x\|_1 = \sum_{t \in \mathbb{N}} |x_t|$. For $x \in \ell_\infty$ and $m \in \ell_1$, we use the notation $m \cdot x = x \cdot m = \sum_{t=0}^\infty m_t x_t$. Countably additive probability measures on \mathbb{N} are identified with the corresponding positive sequences $m \in \ell_1$. Then $x \cdot m$ denotes the expectation of $x \in \ell_\infty$ with respect to the positive measure m .

The sequence $(1, 1, \dots)$, which is identically 1, is denoted by $\mathbf{1}$.

Let \succeq be a weak order, and let

- $\mathcal{U} = \{x \in \ell_\infty : x \succeq 0\}$, and
- $\mathcal{P}_D = \{x \in \ell_\infty : \forall \delta \in D, (1 - \delta) \sum_{t=0}^\infty \delta^t x_t \geq 0\}$.

Let \succeq satisfy D -MON, CONT, COC, and CVX.

LEMMA 4: $x \succeq y$ implies $x \succeq y$, and $x \gg y$ implies $x \succ y$.

Lemma 4 is trivial. Observe that, as a consequence of Lemma 4, $\mathcal{U} \neq \ell_\infty$.

LEMMA 5: \mathcal{U} and \mathcal{P}_D are closed convex cones, and $\mathcal{P}_D \subseteq \mathcal{U}$.

PROOF: That \mathcal{U} is closed follows from continuity of \succeq , convexity follows from CVX, and that \mathcal{U} is a cone is a consequence of \succeq satisfying COC. The corresponding properties for \mathcal{P}_D are straightforward to verify. That $\mathcal{P}_D \subseteq \mathcal{U}$ follows from D -monotonicity. *Q.E.D.*

LEMMA 6: There is $M \subseteq \ell_1$, convex and weakly compact, such that $m \geq 0$ for all $m \in M$, and such that

$$\mathcal{U} = \bigcap_{m \in M} \{x \in \ell_\infty : m \cdot x \geq 0\}.$$

PROOF: The argument is similar to Theorem 1 of Chateauneuf, Maccheroni, Marinacci, and Tallon (2005), observing that monotone continuity is implied by our notion of continuity, and that the variation norm on the set of measures coincides with what we are calling the ℓ_1 norm. We sketch it here for completeness.

Step 1: Constructing a set M of finitely additive probabilities on \mathbb{N} as the polar cone of \mathcal{U} .

Let $\text{ba}(\mathbb{N})$ denote the bounded, additive set functions on \mathbb{N} , and observe that $(\ell_\infty, (\text{ba}(\mathbb{N})))$ is a dual pair. Consider the cone $M^* \subseteq \text{ba}(\mathbb{N})$ given by $M^* = \bigcap_{x \in \mathcal{U}} \{m : m(x) \geq 0\}$. By Aliprantis and Border (1999), Theorems 5.86 and 5.91, $\mathcal{U} = \bigcap_{m \in M^*} \{x : m(x) \geq 0\}$.¹⁰ Since $\ell_\infty^+ \subseteq \mathcal{U}$ (by monotonicity of \succeq), we can conclude that $M^* \subseteq \text{ba}(\mathbb{N})^+$. Moreover, there is nonzero $m \in M^*$ (as $\mathcal{U} \neq \ell_\infty$). For any such nonzero m , observe that since $m \geq 0$, it follows that $m(\mathbf{1}) > 0$.¹¹ Let $M = \{m \in M^* : m(\mathbf{1}) = 1\}$ and conclude that $\mathcal{U} = \bigcap_{m \in M} \{x : m \cdot x \geq 0\}$.

¹⁰One needs to verify that \mathcal{U} is weakly closed with respect to the pairing $(\ell_\infty, \text{ba}(\mathbb{N}))$, but it is by Theorem 5.86 since $\text{ba}(\mathbb{N})$ are the ℓ_∞ continuous linear functionals by Aliprantis and Border (1999, Theorem 12.28).

¹¹Otherwise, we would have $m(x) = 0$ for all $x \in [0, 1]$, which would imply that $m = 0$.

Step 2: Verifying that all elements of M are countably additive.

We show now that each $m \in M$ is countably additive. Since, for all $\theta \in [0, 1)$, there is T so that

$$\underbrace{(1 - \theta, \dots, 1 - \theta, -\theta, -\theta, \dots)}_{T \text{ times}} \in \mathcal{U}$$

(by monotonicity and continuity), it follows that for all $m \in M$, $m(\{0, \dots, T - 1\}) \geq \theta$. Conclude that $\lim_{t \rightarrow \infty} m(\{0, \dots, t\}) = m(\mathbf{N})$, so that countable additivity is satisfied.¹² So we write $m(z) = m \cdot z$.

Step 3: Establishing that M is weakly compact.

Countably additive and nonnegative set functions can be identified with elements of ℓ_1 , so we can view M as a subset of ℓ_1 . We show that M is weakly compact, under the pairing (ℓ_1, ℓ_∞) .

We first show that M is *tight* as a collection of measures: for all $\varepsilon > 0$, there is a compact (finite) set $E \subseteq \mathbf{N}$ such that $m(E) > 1 - \varepsilon$ for all $m \in M$. So let $\varepsilon > 0$ and $\theta' \in (1 - \varepsilon, 1)$. Then, by monotonicity, co-cardinality, and continuity, we know that there is T such that

$$\underbrace{(1 - \theta', \dots, 1 - \theta', -\theta', \dots)}_{T \text{ times}} \in \mathcal{U}.$$

The set $E = \{0, \dots, T - 1\}$ works in the definition of tightness because for every $m \in M$, we have $m(\{0, \dots, T - 1\}) \geq \theta' > 1 - \varepsilon$.

The weak compactness of M then follows from a few simple identifications. Denote the set of countably additive probability measures on \mathbf{N} by $\mathcal{P}(\mathbf{N})$, and the set of nonnegative summable sequences which sum to 1 by $\mathbf{1}(\mathbf{N})$. Observe that the weak* topology on $\mathcal{P}(\mathbf{N})$ induced by the pairing $(\ell_\infty, \mathcal{P}(\mathbf{N}))$ coincides with the weak topology on $\mathbf{1}(\mathbf{N})$ induced by the pairing $(\mathbf{1}(\mathbf{N}), \ell_\infty)$, when in the second instance we identify each $m \in \mathcal{P}(\mathbf{N})$ with an element of $\mathbf{1}(\mathbf{N})$. By Lemma 14.21 of [Aliprantis and Border \(1999\)](#), since M is tight, its closure is compact in the first topology (and hence the second). But M is already closed, as the intersection of a collection of closed sets.¹³ Therefore, we know that every net in M has a subnet which converges in the weak topology on $\mathbf{1}(\mathbf{N})$. Viewing now M as a subset of ℓ_1 , we know that every net in M has a convergent subnet in the weak topology induced by the pairing (ℓ_1, ℓ_∞) , which is what we wanted to show. *Q.E.D.*

LEMMA 7: *The function $U : \ell_\infty \rightarrow \mathbf{R}$ defined by $U(x) = \min\{m \cdot x : m \in M\}$ represents \succeq .*

PROOF: We first establish that $U(x) = 0$ if and only if $x \sim 0$. To see this, suppose $U(x) = 0$. Then, by definition of U , $x \in \mathcal{U}$; thus $x \geq 0$. To establish that $x \sim 0$, we rule out that $x > 0$. So suppose that $x > 0$. Then we would have, by continuity of \succeq , that there is $\epsilon > 0$ small so that $x \geq \epsilon \mathbf{1}$. But then $x - \epsilon \mathbf{1} \in \mathcal{U}$ (by COC of \succeq), so that $U(x - \epsilon \mathbf{1}) \geq 0$. Then the definition of U implies that $U(x) \geq \epsilon > 0$, a contradiction. So $U(x) = 0$ implies $x \sim 0$.

¹²For example, see [Aliprantis and Border \(1999, Lemma 9.9\)](#). Suppose $E_k \subset \mathbf{N}$ is a sequence of sets for which $\bigcap_k E_k = \emptyset$ and $E_{k+1} \subseteq E_k$. Then for each k , there is $t(k) \in \mathbf{N}$ such that $E_k \subseteq \{t(k), t(k) + 1, \dots\}$ and for which $t(k) \rightarrow \infty$. Without loss, take t to be nondecreasing. The result then follows as $m(E^k) \leq m(\{t(k), t(k) + 1, \dots\}) \rightarrow 0$.

¹³Namely, the sets $\{m : p \cdot m \geq 0\}$ for $p \in P$ and $\{m : \mathbf{1} \cdot p = 1\}$.

Conversely, suppose that $x \sim 0$. It follows that $x \in \mathcal{U}$, from which we obtain $U(x) \geq 0$. If in fact $U(x) > 0$, then let $\varepsilon > 0$ be such that $U(x) \geq \varepsilon$, and hence $U(x - \varepsilon \mathbf{1}) \geq 0$. Thus $x - \varepsilon \mathbf{1} \in \mathcal{U}$, so that $x \succ x - \varepsilon \mathbf{1} \geq 0$, or $x \succ 0$, a contradiction.

So now let $x \in \ell_\infty$ be arbitrary. We claim that $x \sim U(x)\mathbf{1}$. But this follows directly from COC as $U(x - U(x)\mathbf{1}) = 0$ if and only if $x - U(x)\mathbf{1} \sim 0$ if and only if $x \sim U(x)\mathbf{1}$.

Finally, the result follows a standard textbook argument: if $x \geq y$, $x \sim U(x)\mathbf{1}$, and $y \sim U(y)\mathbf{1}$, it must be that $U(x) \geq U(y)$; otherwise we would have $U(y)\mathbf{1} \gg U(x)\mathbf{1}$, and hence $U(y)\mathbf{1} \succ U(x)\mathbf{1}$ by monotonicity. Conversely, if $U(x) \geq U(y)$, we have by monotonicity that $U(x)\mathbf{1} \geq U(y)\mathbf{1}$, so that $x \geq y$. Q.E.D.

LEMMA 8: *The function $m : [0, 1) \rightarrow \ell_1$ given by $m(\delta) = (1 - \delta)(1, \delta, \delta^2, \dots)$ is norm-continuous.*

PROOF: First, we show that the map $d : [0, 1) \rightarrow \ell_1$ given by $d(\delta) = (1, \delta, \delta^2, \dots)$ is continuous. The result will then follow as $m(\delta) = (1 - \delta)d(\delta)$.¹⁴

So, let $\delta_n \rightarrow \delta^*$. Then $\|d(\delta_n) - d(\delta^*)\|_1 = \sum_t |\delta_n^t - (\delta^*)^t|$. Observe that for each t , $|\delta_n^t - (\delta^*)^t| \rightarrow 0$. By letting $\hat{\delta} = \sup_n(\delta_n) < 1$, we have that for each t , $|\delta_n^t - (\delta^*)^t| \leq \max\{ |(\delta^*)^t|, |\hat{\delta}^t - (\delta^*)^t| \}$, since the expression $|\delta^t - (\delta^*)^t|$ increases monotonically when δ moves away from δ^* . And observe that $\sum_t \max\{ |(\delta^*)^t|, |\hat{\delta}^t - (\delta^*)^t| \} < +\infty$. Conclude by the Lebesgue Dominated Convergence Theorem (Theorem 11.20 of Aliprantis and Border (1999)) that $\|d(\delta_n) - d(\delta^*)\|_1 \rightarrow 0$. Q.E.D.

LEMMA 9: *Let $D \subseteq (0, 1)$. Then*

$$m((0, 1)) \cap \overline{\text{co}}(m(D)) \subseteq m(D).$$

PROOF: Let $\delta \in (0, 1)$ and suppose that $m(\delta) \in \overline{\text{co}}(m(D))$. Then there is a sequence $m^n \in \text{co}(m(D))$ such that $m^n \rightarrow m$. For each n , $m^n = \sum_i \lambda_i^n m(\delta_i^n)$ for some λ_i^n, δ_i^n . View each λ^n as a probability measure on D ; thus $m^n = \mathbf{E}_{\lambda^n} m(\delta)$. The set of probability measures on D is weak*-compact (Theorem 6.25 of Aliprantis and Border (1999)), so there is a probability measure μ on D so that (taking a subsequence if necessary) $\lambda^n \rightarrow_{w^*} \mu$. This implies that for each t ,

$$m_t^n \rightarrow \mathbf{E}_\mu m_t(\delta) = \mathbf{E}_\mu (1 - \delta) \delta^t$$

by definition. But weak convergence of m^n implies pointwise convergence. Thus, $m_t = \mathbf{E}_\mu (1 - \delta) \delta^t$ by the uniqueness of limits.

If $m(\delta)$ is an extreme point of $\overline{\text{co}}(m(D))$, then μ must be degenerate and therefore $\delta \in D$. So all the extreme points of $\overline{\text{co}}(m(D))$ are in $m(D)$. If $m(\delta)$ were not an extreme point, then it would be a convex combination of elements of $m(D)$, but this is not possible as all the elements of $m((0, 1))$ are extreme points. Q.E.D.

LEMMA 10: *M is the weakly closed convex hull of its strongly exposed points.*

PROOF: By Lemma 6, M is weakly compact. The result follows from Corollary 5.18 of Benyamini and Lindenstrauss (1998). Q.E.D.

¹⁴The latter is easily deemed continuous. By a simple application of the triangle inequality, if $\delta_n \rightarrow \delta^*$, we have $\|(1 - \delta_n)d(\delta_n) - (1 - \delta)d(\delta)\|_1 \leq |(\delta - \delta_n)|\|d(\delta_n)\|_1 + (1 - \delta)\|d(\delta_n) - d(\delta)\|_1$.

LEMMA 11—Theorem 1.1 of [Shohat and Tamarkin \(1943\)](#): Let $\gamma : \mathbf{Z}_+ \rightarrow \mathbf{R}$ with $\gamma(0) = 1$. Let $F \subseteq \mathbf{R}$ be closed. There is a Borel probability measure η on F for which $\gamma(t) = \int_F \delta^t d\eta(\delta)$ iff, for every $n \in \mathbf{N}$, and every polynomial $P(\delta) = \sum_{i=0}^n a_i \delta^i$ on \mathbf{R} ,

$$\forall \delta \in F, \quad P(\delta) \geq 0 \implies \sum_{i=0}^n a_i \gamma_i \geq 0.$$

REMARK 8: [Shohat and Tamarkin \(1943\)](#) stated the two-dimensional version of this problem, but it is straightforward to see that it implies the statement in Lemma 11. The hypothesis that $\gamma(0) = 1$ ensures that η is a probability measure.

LEMMA 12: For each $m \in M$, there exists a Borel probability measure μ on D for which

$$m_t = \int_D (1 - \delta)^t d\mu(\delta).$$

PROOF: Let $m \in M$. Note that we must have $m_0 > 0$ by monotonicity. Define $\gamma : \mathbf{Z}_+ \rightarrow \mathbf{R}$ with $\gamma(0) = 1$ by $\gamma(t) = m_t/m_0$. Choose a polynomial $P : \mathbf{R} \rightarrow \mathbf{R}$, with $P(\delta) = \sum_{i=0}^n a_i \delta^i$, such that $P(\delta) \geq 0$ for all $\delta \in D$. Then, by definition of \mathcal{P}_D ,

$$(a_0, a_1, \dots, a_n, 0, \dots) \in \mathcal{P}_D.$$

Since $\mathcal{P}_D \subseteq \mathcal{U}$, $\sum_{t=0}^{\infty} m_t a_t \geq 0$. Thus $\sum_{t=0}^{\infty} \gamma(t) a_t \geq 0$.

By Lemma 11, there exists a Borel probability measure η on D for which $\gamma(t) = \int_D \delta^t d\eta(\delta)$. Define a probability measure μ on D by

$$\mu(A) = \frac{1}{\int_D (1 - \delta)^{-1} d\eta(\delta)} \int_A (1 - \delta)^{-1} d\eta(\delta)$$

for every Borel set $A \subseteq D$. The measure μ is well defined as D is bounded away from 1; the Radon–Nikodym derivative $d\mu/d\eta$ is

$$\frac{(1 - \delta)^{-1}}{\int_D (1 - \delta)^{-1} d\eta(\delta)}$$

on D .

Now,

$$\sum_{t=0}^{\infty} \gamma(t) = \sum_{t=0}^{\infty} \int_D \delta^t d\eta(\delta) = \int_D \frac{1}{(1 - \delta)} d\eta(\delta)$$

by Fubini's theorem (Theorem 11.26 of [Aliprantis and Border \(1999\)](#)).

Finally, observe that

$$\int_D (1 - \delta) \delta^t d\mu(\delta) = \frac{\int_D \delta^t d\eta(\delta)}{\int_D (1 - \delta)^{-1} d\eta(\delta)} = \frac{\gamma(t)}{\sum_{s=0}^{\infty} \gamma(s)} = m_t.$$

Q.E.D.

4.1. Proof of Theorem 1

We omit the straightforward proof of necessity of the axioms.

For each $m \in M$, let μ_m be the Borel probability measure on D defined by Lemma 12. Let $\Sigma = \{\mu_m : m \in M\}$.

Let $x \in \ell_\infty$. By Lemma 8, $\delta \mapsto \sum_{t=0}^{\infty} (1 - \delta) \delta^t x_t$ is continuous. By Fubini's theorem (Theorem 11.26 of Aliprantis and Border (1999)),

$$\sum_{t=0}^{\infty} \left(\int_D (1 - \delta) \delta^t d\mu(\delta) \right) x_t = \int_D \left(\sum_{t=0}^{\infty} (1 - \delta) \delta^t x_t \right) d\mu(\delta).$$

Further,

$$\mu \mapsto \int_D \left(\sum_{t=0}^{\infty} (1 - \delta) \delta^t x_t \right) d\mu(\delta)$$

is continuous in the weak* topology on $\Delta(D)$.

Then

$$\begin{aligned} U(x) &= \min\{m \cdot x : m \in M\} = \min \left\{ \sum_{t=0}^{\infty} \left(\int_D (1 - \delta) \delta^t d\mu_m(\delta) \right) x_t : m \in M \right\} \\ &= \min_{\mu \in \Sigma} \sum_{t=0}^{\infty} \left(\int_D (1 - \delta) \delta^t d\mu(\delta) \right) x_t. \end{aligned}$$

The theorem follows from Lemma 7.

4.2. Proof of Theorem 2

We omit the straightforward proof of necessity of the axioms.

Suppose that \succeq satisfies IOU, in addition to the axioms in Theorem 1. Then \succeq has a multi-utilitarian representation, obtained from $U(x) = \min\{m \cdot x : m \in M\}$ as in the proof of Theorem 1. We shall prove that M is a singleton.

Suppose by means of contradiction that there are $m, m' \in M$ for which $m \neq m'$. Let $x \in \mathcal{U}$ be such that $m \cdot x < 0 < m' \cdot x$.¹⁵ By Lemma 6, $m \cdot x^* < 0$ and $m' \cdot (-x^*) < 0$, $x \notin \mathcal{U}$ and $-x \notin \mathcal{U}$. Then completeness implies that $0 \succ x$, and $0 \succ -x$. But by applying IOU, we observe $0 - x \succeq x - x$, or $-x \succeq 0$, a contradiction to $0 \succ -x$.

4.3. Proof of Theorem 3

We omit the straightforward proof of necessity of the axioms.

Suppose that \succeq satisfies ISTAT and COMP, in addition to the axioms in Theorem 1. Then \succeq has a multi-utilitarian representation, obtained from $U(x) = \min\{m \cdot x : m \in M\}$ as in the proof of Theorem 1. We shall prove that M is the closed convex hull of sequences of the form $m(\delta)$, for $\delta \in D$. The main idea in the proof is to work with the *exposed* points of M (using Lemma 10) and show that ISTAT implies that they have the memory-less property that characterizes the geometric distribution.

¹⁵For example, let $x^* \in \mathcal{U}$ be such that $m \cdot x^* < m' \cdot x^*$, let $\theta \in (m \cdot x^*, m' \cdot x^*)$, and take $x = x^* - \theta$.

We first observe that, by COMP, all of the measures $m \in M$ satisfy the property that for all T , $m(\{T, T+1, \dots\}) > 0$.

Now, recall that a point of M is *exposed* if there is a continuous linear functional f with $f(m) < f(m')$ for all $m' \in M \setminus \{m\}$. We now show that any exposed point of M has the form $(1-\delta)(1, \delta, \delta^2, \dots)$ for some $\delta \in [0, 1]$.

So, suppose that $m \in M$ is an exposed point. Then there exists $x \in \ell_\infty$ such that $x \cdot m < x \cdot m'$ for all $m' \in M \setminus \{m\}$. Since $x \in \mathcal{U}$, $x \cdot m \geq 0$. In fact, $x \cdot m = 0$ because if $x \cdot m > 0$, observe that $x - (x \cdot m)\mathbf{1}$ satisfies $0 = (x - x \cdot m\mathbf{1}) \cdot m < (x - x \cdot m\mathbf{1}) \cdot m'$ for all $m' \in M \setminus \{m\}$. From $x \cdot m = 0$, it follows that x is on the boundary of \mathcal{U} .

We now show that since $x \in \text{bd}(\mathcal{U})$, we have $(0, \dots, 0, x) \in \mathcal{U}$ and $x + (0, \dots, 0, x) \in \text{bd}(\mathcal{U})$. Observe first that, by continuity and monotonicity, $x \in \text{bd}(\mathcal{U})$ if and only if $x \sim 0$: If $x \sim 0$, then for any ϵ , $x + \epsilon\mathbf{1} \succ x$ and $x \succ x - \epsilon\mathbf{1}$, so $x \in \text{bd}(\mathcal{U})$. On the other hand, if $x \in \text{bd}(\mathcal{U})$, then any open ball about x intersects both $\{y : y \succ 0\}$ and $\{y : 0 \succ y\}$, so it follows by continuity that $x \sim 0$.

So let $x \in \text{bd}(\mathcal{U})$. Then $x \sim 0$. Therefore, indifference stationarity implies that $(0, 0, \dots, 0, x) \sim 0$, and that $(1/2)x + (1/2)(0, \dots, 0, x) \sim 0$. Thus $(0, \dots, 0, x) \in \mathcal{U}$, and, co-cardinality implies that $x + (0, \dots, 0, x) \sim 0$, so $x + (0, \dots, 0, x) \in \text{bd}(\mathcal{U})$.

Since $x + \underbrace{(0, \dots, 0, x)}_{T \text{ times}}$ is on the boundary of \mathcal{U} , there is $m^x \in M$ for which, for all $y \in \mathcal{U}$,

$$0 = m^x \cdot \left(x + \underbrace{(0, \dots, 0, x)}_{T \text{ times}}\right) \leq m^x \cdot y. \quad {}^{16}$$

But observe that, since $x \in \mathcal{U}$ and $\underbrace{(0, \dots, 0, x)}_{T \text{ times}} \in \mathcal{U}$, $m^x \cdot x \geq 0$ and $m^x \cdot (0, \dots, 0, x) \geq 0$.

Then $0 = m^x \cdot \underbrace{(0, \dots, 0, x)}_{T \text{ times}} + m^x \cdot x$ means that $m^x \cdot x = 0$ and $m^x \cdot (0, \dots, 0, x) = 0$. But

$m^x \cdot x = 0$ implies that $m^x = m$, as x was chosen to expose m . In turn, $m^x = m$ implies that $m \cdot \underbrace{(0, \dots, 0, x)}_{T \text{ times}} = 0$ as well.

Let

$$m^T = \frac{(m(T-1), m(T), m(T+1), \dots)}{m(\{T-1, \dots\})} \in \ell_1$$

(recall that we established that $m(\{T-1, \dots\}) > 0$). We shall first show that $m^T \in M$. To this effect, let $p \in \mathcal{U}$ be arbitrary. It is enough to show that $\underbrace{(0, \dots, 0, p)}_{T \text{ times}} \in \mathcal{U}$, as $m^T \cdot$

$p = m \cdot (0, \dots, 0, p) \geq 0$ and $p \in \mathcal{U}$ is arbitrary. So let $0 \leq c = \inf\{p \cdot m' : m' \in M\}$, and note that $0 = \inf\{(p - c\mathbf{1}) \cdot m' : m' \in M\}$, the infimum being achieved at some $m' \in M$ by compactness of M . Then $p - c\mathbf{1} \in \text{bd}(\mathcal{U})$.

Now, $p - c\mathbf{1} \sim 0$ and indifference stationarity imply that $(0, \dots, 0, p - c\mathbf{1}) \in \mathcal{U}$. Then, by monotonicity, we have that $(0, \dots, 0, p) \in \mathcal{U}$. This establishes that $m^T \in M$.

Now, $m^T \cdot x = 0$ and x exposes m , so $m^T \in M$ implies that $m = m^T$. This equation ($m^T = m$ for all T) characterizes the geometric distribution: Let $h(s) = m(\{s, s+1, \dots\})$.

¹⁶There is a supporting hyperplane (Aliprantis and Border (1999, Lemma 5.78)) which we can identify with $m^x \in M^*$ (using the notation from the proof of Lemma 6). In fact, by definition of supporting hyperplane, the linear functional in question is nonconstant, and therefore we have $m^x \in M$. That the supporting hyperplane passes through zero follows as \mathcal{U} is a cone. That m^x is in the polar cone to \mathcal{U} follows by definition.

Then we have

$$\begin{aligned}\frac{h(s+t)}{h(t)} &= \frac{m(\{t+s, t+s+1, \dots\})}{m(\{t, t+1, \dots\})} \\ &= m(\{s, s+1, \dots\}) = h(s).\end{aligned}$$

Then we obtain $h(t) = h((t-1)+1) = h(t-1)h(1)$. Continuing by induction $h(t) = h(1)^t$. If we let $\delta = h(1) = m^*(\{1, 2, \dots\})$, we have $m^*(\{t, \dots\}) = \delta^t$ for all $t \geq 1$, and $m^*(\{0\}) = 1 - m^*(\{1, \dots\}) = 1 - \delta$. Finally, observe $\delta > 0$ as $m(\{T, \dots\}) > 0$ for all T .

So, conclude that each exposed point of M takes the form $(1-\delta)(1, \delta, \delta^2, \dots)$ for some $\delta > 0$ (and clearly $\delta < 1$).

We proceed to proving the theorem. By Lemma 7, $U(x) = \min\{m \cdot x : m \in M\}$ represents \succeq . The previous argument and Lemma 10 show that there exists some $D' \subseteq (0, 1)$ such that M is the weakly closed convex hull of

$$m(D') = \{(1-\delta)(1, \delta, \delta^2, \dots) : \delta \in D'\}.$$

It remains to show that $D' \subseteq D$. So let $\delta \in (0, 1) \setminus D$. By Lemma 9, $m(\delta) \notin \overline{\text{co}}(m(D))$. Then the separating hyperplane theorem implies the existence of $x \in \ell_\infty$ such that

$$m(\delta) \cdot x < 0 \leq \inf\{m(\tilde{\delta}) \cdot x : \tilde{\delta} \in D\}.$$

But then $\sum_t (1 - \tilde{\delta}) \tilde{\delta}^t x_t \geq 0$ for all $\tilde{\delta} \in D$. So $x \geq 0$ by D -MON. Hence $\delta \notin D'$, as U represents \succeq and $x \geq 0$.

The necessity of the axioms is straightforward and omitted. We only provide the calculations showing that the representation satisfies stationarity. Let $x \sim \theta$, so $\theta = U(x) = \min_{\delta \in D} (1-\delta) \sum_t \delta^t x_t$, where the minimum is achieved for some $\delta \in D$.

Let $z = \lambda x + (1-\lambda)(\theta, \dots, \theta, x)$. Then, for any δ ,

$$(1-\delta) \sum_t \delta^t z_t = \lambda(1-\delta^T)\theta + [\lambda + (1-\lambda)\delta^T](1-\delta) \sum_t \delta^t x_t.$$

But $\theta = \lambda(1-\delta^T)\theta + [\lambda + (1-\lambda)\delta^T]\theta$, so for $\delta \in D$, $(1-\delta) \sum_t \delta^t z_t \geq \theta$ if and only if $(1-\delta) \sum_t \delta^t x_t \geq \theta$. A similar statement holds for equalities. This implies that $U(z) = \theta$.

REFERENCES

- ALIPRANTIS, C. D., AND K. C. BORDER (1999): *Infinite Dimensional Analysis: A Hitchhiker's Guide* (Second Ed.). Berlin: Springer. [1330,1338-1343]
- ALON, S., AND G. GAYER (2016): "Utilitarian Preferences With Multiple Priors," *Econometrica*, 84 (3), 1181–1201. [1336]
- ARROW, K. J. (1971): "The Theory of Risk Aversion," in *Essays in the Theory of Risk-Bearing*. Amsterdam: North-Holland, 90–120. [1329]
- BASTIANELLO, L., AND A. CHATEAUNEUF (2016): "About Delay Aversion," *Journal of Mathematical Economics*, 63, 62–77. [1335,1337]
- BASU, K., AND T. MITRA (2003): "Aggregating Infinite Utility Streams With Intergenerational Equity: The Impossibility of Being Paretian," *Econometrica*, 71 (5), 1557–1563. [1327]
- BENYAMINI, Y., AND J. LINDENSTRAUSS (1998): *Geometric Nonlinear Functional Analysis*, Vol. 1. American Mathematical Society Colloquium Publications, Vol. 48. Providence, RI: American Mathematical Society. [1340]
- BEWLEY, T. F. (2002): "Knightian Decision Theory. Part I," *Decisions in Economics and Finance*, 25 (2), 79–110. [1336]

- BLACKORBY, C., AND W. BOSSERT (2008): "Interpersonal Comparisons of Well-Being," in *The Oxford Handbook of Political Economy*, ed. by D. A. Wittman and B. R. Weingast. Oxford: Oxford University Press. [1332]
- BLACKORBY, C., W. BOSSERT, AND D. DONALDSON (1994): "Generalized Gini and Cooperative Bargaining Solutions," *Econometrica*, 62 (5), 1161–1178. [1335]
- (1996): "Consistency, Replication Invariance, and Generalized Gini Bargaining Solutions," *Journal of Economic Theory*, 69 (2), 367–386. [1335]
- BØHREN, Ø., AND T. HANSEN (1980): "Capital Budgeting With Unspecified Discount Rates," *Scandinavian Journal of Economics*, 82, 45–58. [1335]
- CHAMBERS, C. P. (2005): "Multi-Utilitarianism in Two-Agent Quasilinear Social Choice," *International Journal of Game Theory*, 33 (3), 315–334. [1335]
- CHAMBERS, C. P., AND F. ECHENIQUE (2016): "On Multiple Discount Rates," Caltech Social Science Working Paper 1418. [1335]
- CHATEAUNEUF, A., F. MACCHERONI, M. MARINACCI, AND J.-M. TALLON (2005): "Monotone Continuous Multiple Priors," *Economic Theory*, 26 (4), 973–982. [1338]
- DANAN, E., T. GAJDOS, B. HILL, AND J.-M. TALLON (2016): "Robust Social Decisions," *American Economic Review*, 106 (9), 2407–2425. [1336]
- D'ASPROMONT, C., AND L. GEVERS (1977): "Equity and the Informational Basis of Collective Choice," *Review of Economic Studies*, 44 (2), 199–209. [1326,1331,1332]
- (2002): "Social Welfare Functionals and Interpersonal Comparability," in *Handbook of Social Choice and Welfare*, ed. by K. Arrow, A. Sen, and K. Suzumura. Handbooks in Economics, Vol. 1. Amsterdam: Elsevier, Chapter 10, 459–541. [1331]
- EKERN, S. (1981): "Time Dominance Efficiency Analysis," *Journal of Finance*, 36 (5), 1023–1033. [1335]
- FENG, T., AND S. KE (2017): "Social Discounting and Long-Run Discounting," Report, University of Michigan. [1337]
- FOSTER, J. E., AND T. MITRA (2003): "Ranking Investment Projects," *Economic Theory*, 22 (3), 469–494. [1335]
- GENEST, C., AND J. V. ZIDEK (1986): "Combining Probability Distributions: A Critique and an Annotated Bibliography," *Statistical Science*, 1 (1), 114–135. [1336]
- GILBOA, I. (1989): "Expectation and Variation in Multi-Period Decisions," *Econometrica*, 57 (5), 1153–1169. [1337]
- GILBOA, I., AND D. SCHMEIDLER (1989): "Maxmin Expected Utility With Non-Unique Prior," *Journal of Mathematical Economics*, 18 (2), 141–153. [1330,1335-1337]
- GILBOA, I., D. SAMET, AND D. SCHMEIDLER (2004): "Utilitarian Aggregation of Beliefs and Tastes," *Journal of Political Economy*, 112 (4), 932–938. [1336]
- HARSANYI, J. C. (1955): "Cardinal Welfare, Individualistic Ethics, and Interpersonal Comparisons of Utility," *Journal of Political Economy*, 63 (4), 309–321. [1337]
- HINOJOSA, M. A., A. M. MÁRMOL, AND J. M. ZARZUELO (2008): "Inequality Averse Multi-Utilitarian Bargaining Solutions," *International Journal of Game Theory*, 37 (4), 597–618. [1335]
- HUBER, P. J. (1981): *Robust Statistics*. Berlin: Springer. [1337]
- HYLLAND, A., AND R. ZECKHAUSER (1979): "The Impossibility of Bayesian Group Decision Making With Separate Aggregation of Beliefs and Values," *Econometrica*, 47 (6), 1321–1336. [1336]
- JACKSON, M. O., AND L. YARIV (2015): "Collective Dynamic Choice: The Necessity of Time Inconsistency," *American Economic Journal: Microeconomics*, 7 (4), 150–178. [1327,1335,1337]
- KOOPMANS, T. C. (1960): "Stationary Ordinal Utility and Impatience," *Econometrica*, 28 (2), 287–309. [1336]
- LAUWERS, L. (2010): "Ordering Infinite Utility Streams Comes at the Cost of a Non-Ramsey Set," *Journal of Mathematical Economics*, 46 (1), 32–37. [1327]
- MARINACCI, M. (1998): "An Axiomatic Approach to Complete Patience and Time Invariance," *Journal of Economic Theory*, 83 (1), 105–144. [1333,1337]
- MONGIN, P. (1995): "Consistent Bayesian Aggregation," *Journal of Economic Theory*, 66 (2), 313–351. [1336]
- (2016): "Spurious Unanimity and the Pareto Principle," *Economics and Philosophy*, 32 (3), 511–532. [1336]
- NISHIMURA, H. (2018): "The Transitive Core: Inference of Welfare From Nontransitive Preference Relations," *Theoretical Economics*, 13 (2), 579–606. [1337]
- NORDHAUS, W. D. (2007): "A Review of the 'Stern Review on the Economics of Climate Change'," *Journal of Economic Literature*, 45, 686–702. [1328]
- OK, E. A., AND Y. MASATLIOGLU (2007): "A Theory of (Relative) Discounting," *Journal of Economic Theory*, 137 (1), 214–245. [1337,1338]
- OK, E. A., AND L. ZHOU (2000): "The Choquet Bargaining Solutions," *Games and Economic Behavior*, 33 (2), 249–264. [1335]
- SEN, A. K. (1970): *Collective Choice and Social Welfare*. San Francisco, CA: Holden-Day. [1326,1331]

- (1979): “Interpersonal Comparisons of Welfare,” in *Economics and Human Welfare: Essays in Honor of Tibor Scitovsky*, ed. by M. J. Boskin. New York: Academic Press, 183–201. [1326,1331]
- SHOHAT, J. A., AND J. D. TAMARKIN (1943): *The Problem of Moments*. Providence, RI: American Mathematical Society. [1330,1336,1341]
- SVENSSON, L.-G. (1980): “Equity Among Generations,” *Econometrica*, 48 (5), 1251–1256. [1327]
- TRANNOY, A., AND T. KARCHER (1999): “Critères de dominance temporelle. Le manuel de l'utilisateur,” *Economie et Prévision*, 2–3 (138–139), 147–162. [1335]
- VARIAN, H. R. (2006): “Recalculating the Costs of Climate Change,” *New York Times*. [1328,1331]
- VILLEGAS, C. (1964): “On Qualitative Probability σ -Algebras,” *The Annals of Mathematical Statistics*, 35, 1787–1796. [1329]
- WAKAI, K. (2008): “A Model of Utility Smoothing,” *Econometrica*, 76 (1), 137–153. [1337]
- WEITZMAN, M. L. (2001): “Gamma Discounting,” *American Economic Review*, 91 (1), 260–271. [1325,1327, 1335,1337]
- ZAME, W. R. (2007): “Can Intergenerational Equity Be Operationalized?” *Theoretical Economics*, 2 (2), 187–202. [1327]
- ZHOU, L. (1997): “Harsanyi’s Utilitarianism Theorems: General Societies,” *Journal of Economic Theory*, 72 (1), 198–207. [1336]
- ZUBER, S. (2011): “The Aggregation of Preferences: Can We Ignore the Past?” *Theory and Decision*, 70 (3), 367–384. [1327,1335,1337]

Co-editor Itzhak Gilboa handled this manuscript.

Manuscript received 17 November, 2016; final version accepted 26 March, 2018; available online 3 April, 2018.